# Does a Peer Recommender Foster Students' Engagement in MOOCs?

Hugues Labarthe, François Bouchet, Rémi Bachelet, Kalina Yacef

# Does a Peer Recommender Foster Students' Engagement in MOOCs?

Hugues Labarthe
LAMOP
University of Paris I
France (75005)
+33 649487203
hugues.labarthe@ac-creteil.fr

François Bouchet
Sorbonne Universités,
UPMC Univ Paris 06
CNRS, LIP6 UMR 7606
75005 Paris, France
+33 144277135
francois.bouchet@lip6.fr

Rémi Bachelet
Centrale Lille
University of Lille
Lille - France
+33 320335466
remi.bachelet@ec-lille.fr

Kalina Yacef
School of Information Technologies
The University of Sydney
Australia
+61 2 9351 6098
kalina.yacef@sydney.edu.au

## ABSTRACT
Overall the social capital of MOOCs is under-exploited. For most students in MOOCs, autonomous learning often means learning alone. Students interested in adding a social dimension to their learning can browse discussion threads, join social medias and may message other students but usually in a blind and somehow random way, only hoping to find someone relevant, available and also willing to interact. This common isolation might be a contributing factor on student attrition rate and on their general learning experience. To foster learners' persistence in MOOCs, we propose to enhance the MOOC experience with a recommender which provides each student with an individual list of rich-potential contacts, created in real-time on the basis of their own profile and activities. This paper describes a controlled study conducted from Sept. to Nov. 2015 during a MOOC on Project Management. A recommender panel was integrated to the experimental users' interface and allowed them to manage contacts, send them an instant message or consult their profile. The population ($N = 8{,}673$) was randomly split into two: a control group, without any recommendations, and an experimental group in which students could choose to activate and use the recommender. After having demonstrated that these populations were similar up to the activation of the recommender, we evaluate the effect of the recommender on the basis of four factors of learners' persistence: attendance, completion, success and participation. Results show the recommender improved all these 4 factors: students were much more likely to persist and engage in the MOOC if they received recommendations than if they did not.

## Keywords
Recommender system, MOOC, persistence, social learning.

## 1. INTRODUCTION
Understanding and reducing the attrition rate in Massive Open Online Courses is still a concern for many scientists, measuring and predicting attrition [2, 10], and trying to uncover its factors [6, 8]. There is a common assumption that students doing well by themselves are more likely to get involved in the learning community. But the paradox is that students do not necessarily know how to initiate and have meaningful conversations within this community, may feel shy or inhibited in such crowded places, which results in further isolation.

Therefore, while learning is above all a social undertaking [1], it turns out that most MOOCs students learn on their own. Far behind the connectivist model, transmissive MOOCs have been implementing functionalities such as synchronous or asynchronous discussions [4], peer grading, potential team mates' geolocation, groups, etc. In such systems, students find others to connect with either in a blind manner or through user-defined filters. Most importantly, contacts are initiated by the students themselves, who need to actively search for others. So it remains extremely difficult to find the right person to interact with in a newly-formed and distance learning MOOC community. This feeling of isolation hinders the learning experience and is a major factor of student attrition [7, 11]. Indeed, the size of students' cohorts and the fact that they usually work at home, at various times and pace, cultivates isolation rather than connection with other students for learning [5], a problem already well-noted before the MOOC era and which led to attempts to reinforce the sense of community [3, 9]. Numerous works have emphasized the need to help people socialising, on the basis that social learning might foster persistence. It requires not only helping students to know how to work with others (and thus to plan tasks for students to perform in a cooperative way), but also in the first place to find relevant potential learning mates one would want to interact with.

In this paper, we address this issue: to foster learners' persistence in MOOCs, we have designed, implemented and tested a recommender system. Our recommender provides each student with a list of high-potential social contacts, on the basis of their own profile and activities. We hypothesise that offering integrated personal data-driven recommendations may increase the students' persistence and success in the MOOC. We chose to consider four key categories of indicators of persistence: attendance, completion, scores and participation.

This paper is organized as follows: in section 2, we present the experiment with our peer recommender, its context and design, the different groups of students considered, the data collected and its preprocessing. In section 3, we analyse the differences in terms of persistence between the experimental groups, and in section 4, we check whether these differences are related to our recommender system. We then conclude the paper with a discussion on limits and on some perspectives of future work.

## 2. EXPERIMENT WITH A PEER RECOMMENDER

### 2.1 Context of the experiment

We built a peer recommender system and deployed it during the 6th session of a French Project Management MOOC[1], powered by Unow[2] using a customised version of the Canvas platform [7]. The course lasted 9 weeks, from September to November 2015 and had a total of 24,980 students enrolled. Chronologically, it started with a 4 week long pre-MOOC period (week -3 to -1), where students could perform some self-assessment, introduce themselves on the discussion threads, explore the platform and so on. Then the 4 week-long core part of the MOOC (week 1 to 4 included) took place, with lecture videos, assignments, quizzes and so on. During the remaining 5 weeks (week 5 to 9), students followed their specialisation modules and took their final exam. In parallel to the main MOOC, students could additionally register to two possible streams: (i) an Advanced Certification stream where, in the first four weeks (1 to 4), learners also had to submit three assignments and perform peer-reviews; (ii) a Team track, where students also had to join a team and practice on a real project. The topic of the MOOC being Project Management, this MOOC assumes that learners, in addition to working individually and autonomously to obtain their certification, should also get involved as much as they can in the community. Figure 1 shows the overall MOOC timeline as well as the number of students who reached various checkpoints in the MOOC [e.g. 7716 students took quiz 1 between week 1 (release time) and week 9 (end of the MOOC)].

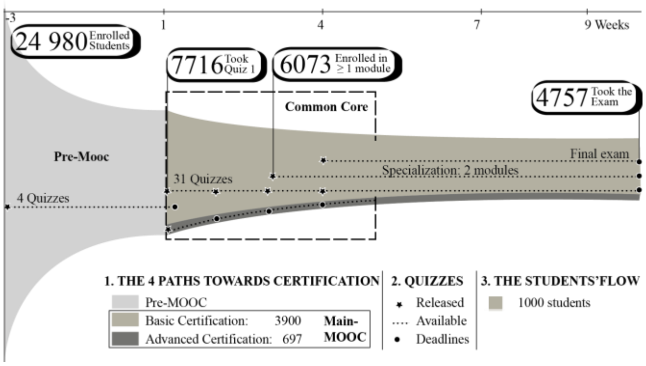


**Figure 1. The 6th edition of the Project Management MOOC: a chronological overview**

### 2.2 The peer recommender widget

The recommendation widget is displayed on the navigation bar on the left side of the screen in a space normally empty (cf. Figure 2). It displays 3 lists: a list of suggested contacts in green, a list of contacts marked as favorite in orange and a list of ignored contacts in grey (A). In each list, other students are represented as a thumbnail showing their name and photo (if any). When bringing the mouse pointer over a thumbnail, it also displays the beginning of their biography (if any) as well as 4 icons: one to send a private message, one to contact them through the chat, one to add them as a favorite and one to ignore them (B). The chat widget is shown on the bottom right-hand corner of the interface and minimised by default. When a message is received, an icon is added and a sound played (C). Bringing the mouse pointer over the widget expands it, giving access to two tabs: in the first tab, the favorite contacts appear and a chat can be initiated with up to 6 of them at the same time. The second tab gives access to a list of previous chats, and one can reopen them to keep interacting with the student(s) associated to that chat (D).

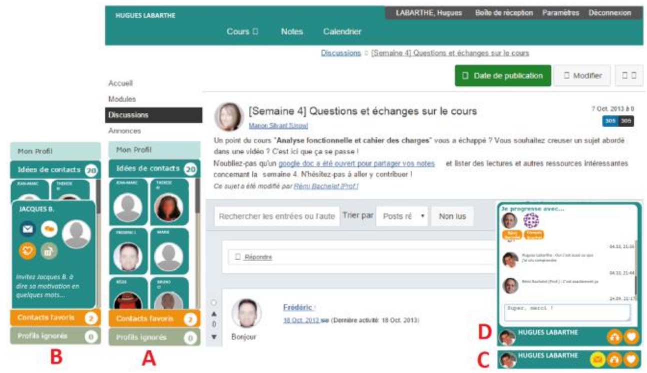


**Figure 2. Recommendations and chat widgets**

### 2.3 Experimental Design

In order to evaluate the effect of the recommender system (RS), we performed a controlled study. A set of experimental groups was offered access to the recommender whilst the control group (***Ctl***) was not. Among the experimental groups, some students accepted the use of the recommender (***ToU***) and others did not. Then among those who accepted it, some interacted with it (***Int***) — i.e. managed contacts, consulted profiles and attempted to write messages— and others did not (***No_Int***) — i.e. had the RS widget visible but did not interact at all with it (an interaction being defined as a click on the interface, as mouse-overs were not recorded). The experimental group was also split in three, each subgroup using a different recommendation algorithm (contact suggestions could be either random, based on social features only, or on a combination of social and advancement features). We shall not compare in this paper the efficiency of these algorithms but focus only on the RS' effect.

### 2.4 Deployment of the Recommender

The recommender was progressively deployed at the beginning of the 4-week core period (week 1 onwards): 100 students on day 1, 4,500 on day 5, 10,000 on day 10. Overall, $N = 8673$ students visiting the platform during this period of time were randomly split between the control group ($N_{Ctl} = 1792$) and the experimental ones ($N_{exp} = 6881$). The experimental group had roughly 3 times more students than the control one because of the aforementioned three subgroups, which will not be considered here. Among students in the experimental groups, $N_{ToU} = 2025$ accepted the recommender Terms of Use (allowing data collection for research purpose) and thus had access to recommendations. Among those students, $N_{Int} = 271$ interacted with the recommendations panel and the chat associated with it (i.e. $N_{No\_Int} = N_{ToU} - N_{Int} = 1754$). Those figures are summarised on Figure 3.

### 2.5 Data Collection and Pre-processing

We extracted two types of data from the MOOC: learning traces as interaction logs, and demographic information coming from students' answers to a demographic questionnaire they could fill during the Pre-MOOC period, or as they started the MOOC for students arriving late on the platform.

One main way to understand how learners behave is by looking at the interaction logs and the learning records. Overall, 3.95 million

[1] MOOC Project Management, http://mooc.gestiondeprojet.pm/

[2] Unow, http://www.unow.fr/

pages were displayed from Sept. 1st to Nov. 22nd (week -3 to 9) for 373,937 different URLs. We classified them into semantic categories consisting of an action and an area of the website. The URLs combine references to 3 main actions: browsing, viewing content, and downloading resources. Students performed these actions on 12 areas as shown in Table 1. In total, students browsed pages with references to 357 different resources: 8.5% are the homepage, 8.3% lesson pages and 43% quizzes. Many students in developing countries download videos on a third-party website, so these figures should only be used to differentiate students' profiles.

We created 10 variables from this learning dataset to capture students' persistence in the MOOC, which could be grouped into four broad categories: attendance, completion, score and participation. These indicators are shown in Table 2.

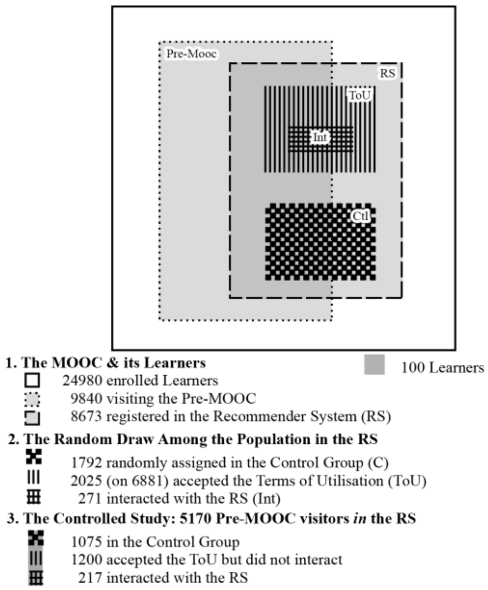


**Figure 3. MOOC cohort sizes and overlaps (to scale)**

**Table 1. Tagging logs towards actions and areas**

| **Categories** • **subcategories** | **Brow-sing** | **View-ing** | **Down-loading** | **Total (%)** |
|---|---|---|---|---|
| [homepage] | 336,941 | | | 8.5 |
| Announcements | 27,768 | | | 0.7 |
| Assignments | 6,602 | 68,591 | | 1.9 |
| • Syllabus | 64,611 | | | 1.6 |
| • Corrected assignments | | 77,270 | | 2.0 |
| • Peer-reviewing materials | | 59,865 | | 1.5 |
| • Downloaded assignments | | 69,510 | 23,606 | 2.4 |
| Calendar | | 2,214 | | 0.1 |
| Discussions | 35,763 | 119,777 | | 3.9 |
| Grades | 42,961 | 27,655 | | 1.8 |
| Modules | 489,325 | | | 12.4 |
| • Badges | | 80,834 | | 2.0 |
| Others | 440 | | | 0.0 |
| Pages | 7,761 | | | 0.2 |
| • Lessons | | 327,882 | | 8.3 |
| • Other Contents | | 323,469 | | 8.2 |
| • Downloads | 58,981 | | | 1.5 |
| Quizzes | 11,713 | 1,686,448 | | 43.0 |
| Profiles | | 2,678 | | 0.1 |
| TOTAL % | 27.4 | 72.0 | 0.6 | 100 |

Finally, in addition to these learning related variables, we extracted the social features from one of three research surveys filled by participants before Nov. 11th. 10,331 learners completed this survey, from which 1,454 were enrolled in the control group and 5,397 in the experimental groups. 6 variables were considered: student's gender, country, year of birth, their level of study (coded as follow: 0, without A-Level; from 1 to 3: years of university course; 4: master degree; 5: PhD), the previous experiences of MOOCs (0 for newcomers, 2 for experienced with MOOCs; 4 for recurring Project Management MOOC students) and the participation to the Pre-MOOC (0 or 1).

**Table 2. Retrieving data related to persistence**

| **Category** | **Indicators** |
|---|---|
| Attendance | 1. Number of days the student visited the platform<br>2. Number of pages the student accessed<br>3. Time spent on these pages [max = 600 s] |
| Comple-tion | 4. Number of attempts to complete a quiz<br>5. Number of quizzes completed |
| Scores | 6. Final score [31 compulsory quizzes + exam] |
| Participa-tion | 7. Number of posts on discussions (forums)<br>8. Average length of discussion posts<br>9. Number of messages sent via the Conversations (private messages)<br>10. Average length of private messages |

## 2.6 Were groups similar before treatment?

In order to assess the similarity between the control group and the experimental ones before the experiment started, we compared their social and behavioral features (cf. Table 3). The data analysis indicates no significant differences between the two groups in terms of gender, countries, year of birth, level of study, previous MOOC experience and attendance on the platform. We can therefore consider the groups were similar before the experiment.

**Table 3. Variation between Groups (ANOVA)**

| **Features (number of values)** | **F** | **P-value** |
|---|---|---|
| Gender (2) | 0.573348 | 0.448958 |
| Countries (91) | 2.14E-06 | 0.998834 |
| Year of Birth (59) | 3.266974 | 0.070732 |
| Level of Study (6) | 1.195992 | 0.274163 |
| Previous experiences of MOOCs (3) | 0.009721 | 0.921462 |
| Participation to the Pre-MOOC (2) | 0.586452 | 0.443815 |

# 3. GROUP BEHAVIOUR ANALYSIS

Table 4 shows the comparison between 3 groups: the control group (*Ctl*), and among the experimental one, the ones which accepted and did (resp. didn't) use the recommender (*No_Int* - resp. *Int*). Figures show the students who experienced RS were those that displayed the strongest values for the 10 indicators of persistence considered. In particular, the average number of daily visits, pages viewed and duration increase from *Ctl* to *No_Int* and *Int*. The standard deviation increases too, revealing that the highest variation of behavior is observed among those who interacted with the RS. In terms of quizzes, the learners who experienced the RS completed 2 more quizzes than the others and scored on average 17 points higher with a smaller standard deviation. Finally, their participation in discussions and conversations are also higher. Reading these figures, it appears that students who experienced the recommender were also more engaged with the course and its community: even though the 271 students in the *Int* group did not spend so much time online overall, they have managed to obtain higher scores in terms of completion, quiz scores and participation.

However, the fact that students who used the recommender were also more engaged is not sufficient to express causality between the two. The uncertainty resides in the fact that in the experimental group, students could *choose whether or not* to have a recommender widget, and *whether or not to* actually make use of it. It could be the case that, in fact, students who are very engaged are more likely to use the recommender.

**Table 4. Average and standard deviation (in italics) of persistence indicators for experimental *versus* control groups**

| | Attendance from W1 to W4 | | | Completion Nov. 22nd | | Scores /100 | Participation from W4 to W9 | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Indicators | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Ctl **N=1792** | 10 *7* | 323 *285* | 1h38 *1h57* | 26.4 *22.5* | 20 *14* | 32.2 *28.7* | 0.7 *3.2* | 69 *137* | 0.3 *2.1* | 31 *127* |
| No_Int **N=1754** | 12 *7.5* | 411 *373* | 2h08 *2h23* | 30.5 *24* | 21.6 *13.3* | 36.1 *30.1* | 1.4 *5.6* | 111 *190* | 0.6 *2.1* | 52 *177* |
| Int **N=271** | 16.1 *6.9* | 616 *405* | 3h46 *3h07* | 43.2 *24.7* | 26.9 *10* | 49.1 *27.8* | 2.7 *6.1* | 154 *186* | 1.6 *3.8* | 107 *212* |

# 4. EFFECTS OF THE RECOMMENDER

To determine the RS' real effect on learners' persistence, we need to compare cohorts that were similar in terms of persistence before the experiment started and see how they evolve during the course of the MOOC. For example, we want to find out whether, among students who were very passive before the recommender was made available, a larger proportion of those who used the recommender persisted in the MOOC. To do so, we first clustered students during the Pre-MOOC period (i.e. before they were allocated to a group, and before the RS was made available) based on their level of engagement (section 4.1). We then, in each cluster, analysed the control and experimental groups according to each dimension of persistence at the end of the main MOOC period.

## 4.1 Pre-MOOC activity clusters

To cluster students in the Pre-MOOC period, we used as features the times spent on 18 of the actions in areas shown in Table 1 (i.e. excluding those related to material not yet available). During the Pre-MOOC, 294,209 pages were accessed by the 9,840 students who were present in the Pre-MOOC period. We used the k-means algorithm to extract clusters and found the best solution involved 4 groups, shown in Table 5 and called A, B, C D on the basis of their time spent (A being the most active and D the least). Students in cluster A spent over 1h40 on the website viewing lessons, quizzes and discussions (sum of the mean values). The second cluster (B) spent less than 40 minutes, essentially in the quizzes area; in the third cluster, C, the time is even shorter and those in the last one, D, stayed less than 2 min on the website in total.

Table 6 shows the distribution across the 4 Pre-MOOC clusters of students who would later belong to groups *Ctl*, *No_Int* and *Int*. Since we want to follow the evolution of the students who were present in the Pre-MOOC period, we must only consider the intersecting population. The populations of the various groups are now: $N_{Pre\&Int}$ = 217 students who interacted with the recommender (vs. $N_{int}$ = 271); $N_{Pre\&No\_Int}$ = 1,200 (vs. $N_{No\_Int}$ = 1,754) who accepted its ToU without using it; $N_{Pre\&Ctl}$ = 1,075 (vs. $N_{Ctl}$ = 1,792) who were randomly enrolled in the control group.

To deal with the sample size difference and compare the features of students in Int with students in *Ctl* and *No_Int*, a subsample was ten times randomly drawn for each cluster – e.g. in the PreMOOC_D cluster, 77 persons out of 551 were ten times randomly drawn. The percentage averages in tables 8, 10 and 12 are computed only on the basis of features of students from these subsamples. We will now exclusively focus on the last 3 Pre-MOOC clusters since the most active group (PreMOOC_A) is very small (8) and already very engaged.

**Table 5. Interactions and clusters during the Pre-MOOC**

| Features (in seconds) | PreMooc _D | PreMooc _C | PreMooc _B | PreMooc _A |
|---|---|---|---|---|
| browsing_homepage | 21 | 48 | 149 | 411 |
| browsing_announcements | 1 | 4 | 15 | 81 |
| browsing_assignment | 4 | 14 | 48 | 210 |
| browsing_discuss._topics | 2 | 8 | 26 | 190 |
| browsing_grades | 1 | 3 | 11 | 30 |
| browsing_modules | 7 | 43 | 140 | 428 |
| browsing_pages | 0 | 1 | 6 | 8 |
| browsing_quizzes | 0 | 1 | 2 | 2 |
| downloading_assignment | 0 | 0 | 0 | 2 |
| viewing_assignment | 1 | 11 | 49 | 208 |
| viewing_calendar_events | 0 | 0 | 0 | 7 |
| viewing_discuss._topics | 13 | 82 | 226 | 857 |
| viewing_grades | 0 | 0 | 1 | 1 |
| viewing_modules | 0 | 7 | 24 | 65 |
| viewing_pages | 25 | 163 | 550 | 1472 |
| viewing_profiles | 0 | 1 | 2 | 37 |
| viewing_quizzes | 33 | 768 | 1167 | 1965 |

**Table 6. Clusters and Groups during the Pre-MOOC**

| | **N (%)** | **N** | **Ctl** | **No_Int** | **Int** |
|---|---|---|---|---|---|
| **PreMooc_D** | 66 | 6,386 | 551 | 578 | 77 |
| **PreMooc_C** | 26 | 2,534 | 393 | 404 | 78 |
| **PreMooc_B** | 7 | 658 | 118 | 190 | 54 |
| **PreMooc_A** | 1 | 62 | 13 | 28 | 8 |
| **Total** | 100 | 9,640 | 1,075 | 1,200 | 217 |

## 4.2 Attendance during the Common Core

We clustered all enrolled students (*N*=24,980) using the full set of features in Table 4 for a total of 3,110,321 pages seen during the Common Core. We obtained 4 clusters, shown in Table 7, named according to their attendance quality (A the best, D the worst). Cluster Att_D, with 77% students, has the poorest overall mean in regards to all the features, not exceeding 6 minutes spent interacting with all pages. The mean values of the second cluster, Att_C (with 17% students), total around 1h30min. The two last clusters, Att_B and Att_A, contain 3% each of the population: the main difference is the time spent by Att_A in the assignments area.

We then explored how the pre-MOOC students evolve into these attendance clusters, according to their activities during the Common Core (cf. Table 8, where figures in a row represent 100% of the mentioned *Ctl*, *No_Int* and *Int*). Considering the lower clusters D to B, these figures suggest that the recommender system played a significant role on the duration of the visits of the learners from clusters D, C and B, that is to say 99% of the Pre-MOOC population. Indeed, one can see that students who used to be in D, having the RS marginally increased their persistence, but significantly increased the persistence of students who used it (32% of them now being in cluster B vs. 8% for students of the control group). For students in clusters C and B during the pre-MOOC, we observe a similar pattern: simply having access to the RS tended to increase their persistence, and actually using the RS tended to significantly decrease their chance of dropping out (i.e. ending up in cluster D, the least active students).

**Table 7. Interactions and clusters during the Common Core**

| Features (in seconds) | At_D | At_C | At_B | At_A |
|---|---|---|---|---|
| _others | 0 | 0 | 1 | 2 |
| browsing_ | 15 | 214 | 554 | 856 |
| browsing_announcements | 1 | 12 | 53 | 61 |
| browsing_assignments | 5 | 48 | 181 | 155 |
| browsing_discussion_topics | 2 | 23 | 90 | 315 |
| browsing_grades | 1 | 32 | 160 | 276 |
| browsing_modules | 22 | 430 | 1022 | 1249 |
| browsing_pages | 0 | 4 | 4 | 6 |
| browsing_quizzes | 0 | 7 | 7 | 6 |
| downloading_assignments | 0 | 3 | 5 | 144 |
| viewing_assignments | 7 | 248 | 636 | 9334 |
| viewing_calendar_events | 0 | 1 | 11 | 5 |
| viewing_discussion_topics | 14 | 127 | 467 | 1477 |
| viewing_grades | 0 | 10 | 48 | 216 |
| viewing_modules | 3 | 57 | 169 | 177 |
| viewing_pages | 67 | 1025 | 2766 | 2398 |
| viewing_profiles | 0 | 1 | 4 | 18 |
| viewing_quizzes | 180 | 3257 | 8286 | 5165 |
| % students | 77 | 17 | 3 | 3 |

**Table 8. Attendance: Evolution of the learners from the Pre-MOOC to the Core-MOOC periods**

| ↓From To→ | At_D | At_C | At_B | At_A | Group |
|---|---|---|---|---|---|
| **PreMooc_D** 66% | 39 | 49 | 8 | 4 | Ctl |
| | 33 | 49 | 12 | 7 | No_Int |
| | 9 | 39 | 32 | 19 | Int |
| **PreMooc_C** 26% | 26 | 50 | 9 | 16 | Ctl |
| | 24 | 43 | 12 | 20 | No_Int |
| | 17 | 45 | 12 | 27 | Int |
| **PreMooc_B** 7% | 16 | 48 | 12 | 24 | Ctl |
| | 16 | 38 | 15 | 31 | No_Int |
| | 2 | 37 | 20 | 41 | Int |

## 4.3 Completion and final scores

We clustered again the student population, using scores and activity in the examination points (i.e. scores obtained at the 31 quizzes and the final exam by the end of the MOOC). Each score is standardised to marks out of 100. We obtained again 4 clusters, which centroids are shown in Table 9. The values of the centroid of the first cluster indicates a large part of students (71%) who participated in the first 2 quizzes but obtained a very low score on them and then did not participate again in any assessment. The centroid of the second cluster (4% of learners) corresponds to students who easily passed the quizzes of the first week but dropped out on the second. The third cluster (4%) has similar students, but who gave up in week 3. Finally, the last cluster (21%) contains all the students who completed all the quizzes and final exam with high scores in each.

Once again figures in Table 10 show that, by accepting the recommendations and, even more, interacting with its panel, the learners went closer to completion and obtained better scores. In particular, we observe as before for students in clusters D and B that the mere presence of the RS has a small positive impact on their chances to complete (or at least to stay longer on the MOOC before giving up), but that students who use the RS benefit the most from an increased chance to complete. For students in cluster C, the use of the RS seems to have made some of them drop out overall a bit later (week 2 instead of week 1) but did not increase their chance to complete the MOOC.

**Table 9. Completion and score clusters during whole MOOC**

| Week | Quiz | D | C | B | A | Week | Quiz | D | C | B | A |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 3 | 92 | 92 | 96 | 2 | 17 | 0 | 1 | 67 | 92 |
| | 2 | 1 | 82 | 82 | 87 | | 18 | 0 | 0 | 48 | 83 |
| | 3 | 0 | 92 | 92 | 96 | | 19 | 0 | 1 | 57 | 95 |
| | 4 | 0 | 82 | 89 | 95 | 3 | 20 | 0 | 1 | 39 | 92 |
| | 5 | 0 | 76 | 93 | 98 | | 21 | 0 | 1 | 40 | 96 |
| | 6 | 0 | 54 | 78 | 87 | | 22 | 0 | 1 | 36 | 95 |
| | 7 | 0 | 63 | 92 | 98 | | 23 | 0 | 1 | 33 | 91 |
| 2 | 8 | 0 | 26 | 93 | 96 | | 24 | 0 | 1 | 31 | 94 |
| | 9 | 0 | 18 | 94 | 97 | | 25 | 0 | 1 | 29 | 89 |
| | 10 | 0 | 10 | 92 | 95 | 4 | 26 | 0 | 1 | 10 | 91 |
| | 11 | 0 | 7 | 88 | 93 | | 27 | 0 | 1 | 5 | 93 |
| | 12 | 0 | 4 | 85 | 93 | | 28 | 0 | 0 | 2 | 90 |
| | 13 | 0 | 2 | 83 | 93 | | 29 | 0 | 1 | 1 | 96 |
| | 14 | 0 | 2 | 86 | 95 | | 30 | 0 | 0 | 1 | 95 |
| | 15 | 0 | 1 | 76 | 89 | | 31 | 0 | 0 | 1 | 86 |
| | 16 | 0 | 1 | 75 | 93 | | EXAM | 0 | 1 | 3 | 78 |
| N (%) | | 71 | 4 | 4 | 21 | N (%) | | 71 | 4 | 4 | 21 |

**Table 10. Completion and final scores: Evolution of the learners from the Pre-MOOC to the Core-MOOC periods**

| ↓From To→ | Co_D | Co_C | Co_B | Co_A | Group |
|---|---|---|---|---|---|
| **PreMooc_D** 66% | 32 | 5 | 13 | 49 | Ctl |
| | 27 | 6 | 14 | 53 | No_Int |
| | 10 | 5 | 4 | 81 | Int |
| **PreMooc_C** 26% | 15 | 9 | 11 | 65 | Ctl |
| | 9 | 9 | 14 | 69 | No_Int |
| | 8 | 14 | 13 | 65 | Int |
| **PreMooc_B** 7% | 8 | 5 | 8 | 79 | Ctl |
| | 5 | 9 | 14 | 73 | No_Int |
| | 4 | 2 | 11 | 83 | Int |

## 4.4 Participation to the Common Core

The total number and average length of the messages sent by each student were retrieved from the Canvas database (discussions and conversations). Using k-means with features from the participation section of Table 2, we obtained once again 4 clusters, shown in Table 11: a first cluster, Pa-D (89% of 24,980 enrolled learners) did not interact at all with others. The centroid of the second one indicates 2 posts of an average of 237 characters on the discussion topics (9%). The third cluster (2%) seems to have a similar activity but slightly stronger in term of number of posts (2.7) and average post length (599 characters). The last 1% is highly committed to the course and its community: most of them correspond to students who were part of the advanced certification stream.

Table 12 shows how students in the Pre-MOOC clusters are distributed over the 4 participation clusters at the end of the MOOC. Figures reveal a consistent positive effect of the mere presence of the RS across the initial Pre-MOOC clusters: there are always less students in cluster Pa_D in the *No_Int* group than in the control group. Less surprisingly, students who interacted with the RS generally did so to send a message to someone, so they overall also ended up less often being in a situation where they do not interact at all with anyone else (complete isolation). Finally, we can see that merely giving students access to a recommender panel does not prevent them from being social-lazy: a majority (82%, 88% 69% respectively in clusters D, C and B) of the students who interacted with the RS did not attempt to directly contact anyone else. These figures are however probably lower than they would be if every student had access to the associated direct chat module, and still better than in the Control group (96%, 91% and 80% respectively in clusters D, C and B) who could only contact others in a blind way through the forum or private messages.

**Table 11. Participation Clusters of all enrolled students**

| Attribute | Pa-D | Pa-C | Pa-B | Pa-A |
|---|---|---|---|---|
| Nb** of discussions | 0 | 2 | 2 | 9 |
| Discussions length* | 2 | 237 | 599 | 264 |
| Nb** of conversations | 0 | 0 | 0 | 7 |
| Conversations length* | 1 | 9 | 19 | 542 |
| N% | 89 | 9 | 2 | 1 |

*: average number of characters; **: number of posts/messages sent

**Table 12. Participation: Evolution of the learners from the Pre-MOOC to the Core-MOOC periods**

| ↓From To→ | Pa_D | Pa_C | Pa_B | Pa_A | Group |
|---|---|---|---|---|---|
| **PreMooc_D** 66% | 78 | 18 | 2 | 2 | Ctl |
| | 67 | 25 | 4 | 4 | No_Int |
| | 47 | 35 | 6 | 12 | Int |
| **PreMooc_C** 26% | 76 | 15 | 4 | 5 | Ctl |
| | 69 | 18 | 4 | 9 | No_Int |
| | 62 | 26 | 4 | 9 | Int |
| **PreMooc_B** 7% | 66 | 14 | 4 | 15 | Ctl |
| | 53 | 25 | 6 | 15 | No_Int |
| | 39 | 30 | 7 | 24 | Int |

## 5. Discussion, conclusion and perspectives

We conducted a controlled study during a Project Management MOOC, in which a recommender panel integrated to the user interface provided suggestions and allowed contact management, instant messaging and profile consultation. Students were randomly split into a control group (without any recommendations), and an experimental group (in which they could activate and use our recommender). The number of the students involved in this experience was relatively high: among 6881 selected students, 2025 accepted the Term of Use of the recommender and 279 accessed its functionalities. We have shown that these populations were similar before the activation of the recommender, and evaluated its effect according to four categories of indicators relative to learners' persistence: attendance, completion, success and participation. Results suggested that our recommender improved these four categories of indicators: students are much more likely to persist and engage in the MOOC if they receive recommendations than if they do not.

The main interest was then to evaluate the effect the recommendations might have played in such increased rates of engagement. To do so, we focused on clustering similar learners according to their activities before the beginning of the course, leading to four groups from the least (D) to the most (A) active students. We analysed the way 3 of these 4 groups (representing 99% of the students) were evolving in terms of attendance, completion and score, participation. We observed overall a significant improvement of students' engagement, not only for those who interacted with the recommendations, but, more largely, for all of those accepted using the recommendation system.

This study presented several limitations: (1) for experimental purposes, we restricted the access to the direct communication tool; (2) since not all students had access to the RS and the chat, the teaching team could not use them for pedagogical activities, which could have boosted the effect of the RS; (3) students in the control group were not asked to accept the RS Terms of Use, since they would not be given access to it – however, while it is thus possible that students who accepted the ToU were more motivated, the analysis presented in section 2.6 shows that students in the control and experimental groups were similar in terms of participation before the beginning of the core MOOC and demographics.. Furthermore, the most significant results were obtained comparing students who interacted vs. those who did not interact with the RS, and these results are not affected.

Overall, this controlled study is highly supporting the idea that recommending learners to learners, in such crowded places as MOOC platforms, is an effective way to get them more involved in terms of attendance, completion, scores and participation. In the future, we intend to look into more details the impact of the different recommendation strategies, and the different ways students interacted with the recommendation system.

## 6. ACKNOWLEDGMENTS

This work was funded by the French Educational Board and by the Human-Centred Technology Cluster of the University of Sydney. We thank Unow for letting us deploy our RS on their MOOC.

## 7. REFERENCES


[1] Bandura, A. 1971. *Social Learning Theory*. General Learning Corporation.

[2] Bouchet, F. and Bachelet, R. 2015. Do MOOC students come back for more? Recurring Students in the GdP MOOC. *Proc. of the European MOOCs Stakeholders Summit 2015* (Mons, Belgium), 174–182.

[3] Croft, N., Dalton, A. and Grant, M. 2010. Overcoming Isolation in Distance Learning: Building a Learning Community through Time and Space. *Journal for Education in the Built Environment*. 5, 1, 27–64.

[4] Ferschke, O., Yang, D., Tomar, G. and Rosé, C.P. 2015. Positive Impact of Collaborative Chat Participation in an edX MOOC. *Proc. of Artificial Intelligence in Education: 17th International Conference, AIED 2015, Madrid, Spain, June, 2015*. Springer. 115–124.

[5] Gütl, C., Rizzardini, R.H., Chang, V. and Morales, M. 2014. Attrition in MOOC: Lessons Learned from Drop-Out Students. *Proc. of Learning Technology for Education in Cloud. MOOC and Big Data*. Springer. 37–48.

[6] Kizilcec, R.F., Piech, C. and Schneider, E. 2013. Deconstructing Disengagement: Analyzing Learner Subpopulations in Massive Open Online Courses. *Proc. of the Third International Conference on Learning Analytics and Knowledge* (New York, NY, USA), 170–179.

[7] Labarthe, H., Bachelet, R., Bouchet, F. and Yacef, K. 2016. Towards increasing completion rates through social interactions with a recommending system. *Proc. of the European MOOCs Stakeholders Summit 2016* (Graz, Austria), 471-480.

[8] Rosé, C.P., Carlson, R., Yang, D., Wen, M., Resnick, L., Goldman, P. and Sherer, J. 2014. Social Factors That Contribute to Attrition in MOOCs. *Proc. of the First ACM Conference on Learning @ Scale Conference* (New York, NY), 197–198.

[9] Rovai, A.P. 2002. Building Sense of Community at a Distance. *The International Review of Research in Open and Distributed Learning*. 3, 1.

[10] Yang, D., Sinha, T., Adamson, D. and Rosé, C.P. 2013. Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. *Proc. of the NIPS Data-Driven Education Workshop*.

[11] Yang, D., Wen, M. and Rosé, C.P. 2014. Peer Influence on Attrition in Massive Open Online Courses. *Proc. of the 7th International Conference on Educational Data Mining* (London, UK), 405–406.